

## Homework 2

### Question 1.

1. Hadoop fs -ls Shakespeare  
Found 4 items  
-rw-rw-rw- 1 training supergroup 1784616 2016-09-12 11:08 shakespeare/comedies  
-rw-rw-rw- 1 training supergroup 1479035 2016-09-12 11:08 shakespeare/histories  
-rw-rw-rw- 1 training supergroup 268140 2016-09-12 11:08 shakespeare/poems  
-rw-rw-rw- 1 training supergroup 1752440 2016-09-12 11:08 shakespeare/tragedies
2. Hadoop fs -cat Shakespeare/poems | head -n 16

## SONNETS

TO THE ONLY BEGETTER OF  
THESE INSUING SONNETS  
MR. W. H. ALL HAPPINESS  
AND THAT ETERNITY  
PROMISED BY  
OUR EVER-LIVING POET WISHETH  
THE WELL-WISHING  
ADVENTURER IN  
SETTING FORTH  
cat: Unable to write to output stream.

3. With nodes N1 and N2 on RA  
N3 and N4 on RB  
N5 and N6 on RC

/filelocation B1 B2 B3 B4 B5  
B1: N1, N3  
B2: N2, N5  
B3: N4, N6  
B4: N1, N3  
B5: N2, N5

### Question 2.

Problem 2:

<u>Map Input</u>	<u>Map output</u>	<u>Reducer Input</u>	<u>Reducer output</u>
map(15)	[(3,15),(5,15)]	(2, [24,30])	(2, 54)
map(21)	[(3,21),(7,21)]	(3, [15, 21, 24, 30])	(3, 90)
map(24)	[(2,24),(3,24)]	(5, [15, 30])	(5, 45)
map(30)	[(2,30),(3,30),(5,30)]	(7, [21, 49])	(7, 70)
map(49)	[(7,49)]		

Question 3.

- ADRIANO: 111
  - Whether: 41
  - Love: 2221
  - Loves: 203
  - The: 25578
  - Whether: 79
  - We: 2922
  - Zodiac: 1
- Hadoop fs -cat wordcounts/part-r-00000 | wc -l
  - 29183
- Sort results based on usage
  - Categorize results based on sentiment analysis word categorizations
- The record reader takes the file input and converts into a key value pair. The key is usually the byte offset or some randomly generated key that typically is not used. The value is in this case a single word from the input (Shakespeare poems).
  - Input ex: (TO THE ONLY BEGETTER OF)
  - Output ex: (0, TO), (3, THE), (7, ONLY)...

Question 4:

- Mapper output
  - [(N, 2.0)]
  - [(n,3.0)]
  - (i,2.0)
  - (d,10.0)
  - (n,3.0)
  - (t,3.0)
  - (b,4.0)
  - (t,4.0)

Reducer input

[N,(2.0)]

[n,(3.0,3.0)]

[b,(4.0)]

[d,(10.0)]

[i ,(2.0)]

[t,(3.0,4.0)]

2. Code please refer to the repo

3. A---3.89139

W----- 4.46401

a-----3.077655

t----3.73326

z-----4.672727